

ML

Регрессия

Per.5: Don't assume that low-level code is necessarily faster than high-level code

МОЁ ХОББИ: ЭКСТРАПОЛИРОВАТЬ

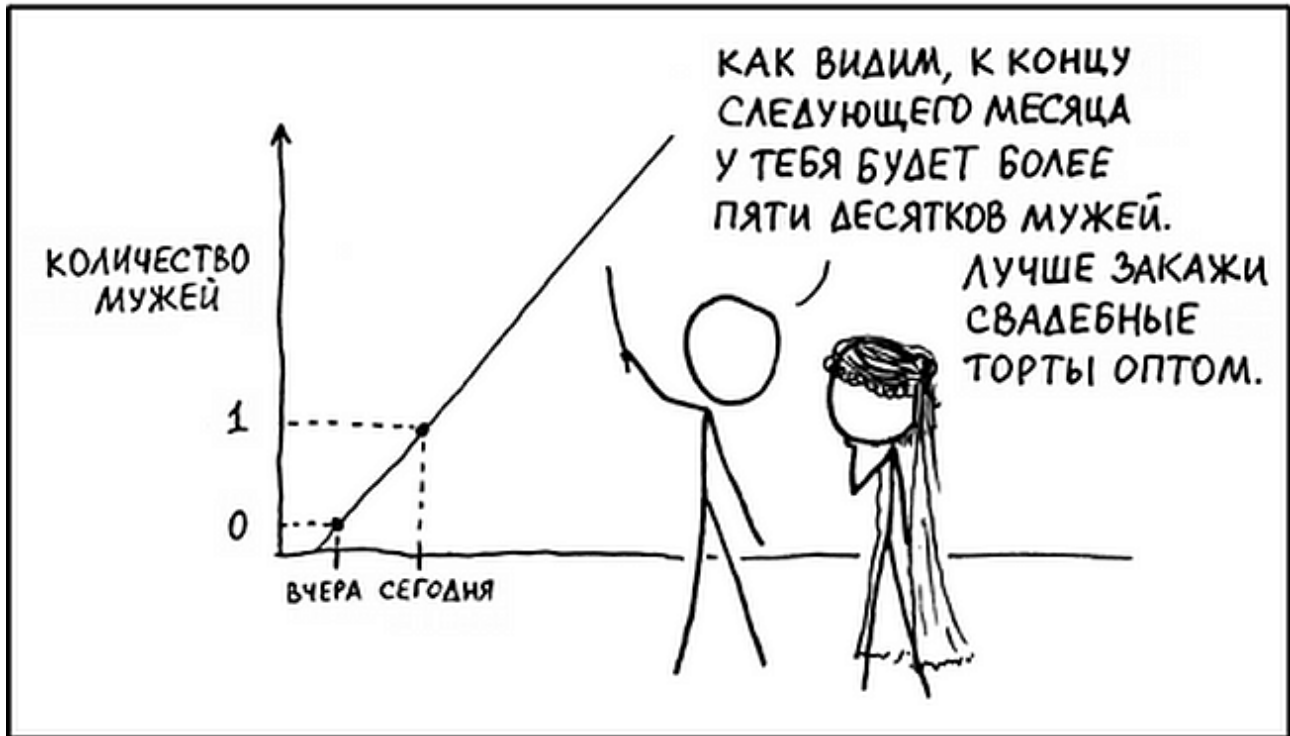


Figure 1: 0_7a848_bb78e424_orig

Сегодня

- Преобразование простого текста
- Бинарная и многоклассовая классификация
- Кластеризация и тематическое моделирование
- Регрессия
- Понижение размерности
- Рекомендации

Преобразование простого текста

How to format my hard disk. Hard disk format problems.

disk format hard how my problems to

X, Y, X X X, Y Y Y X

How to format my hard disk. => 1 1 1 1 1 0 1

disk format hard how my problems to

1 1 1 1 1 0 1

Hard disk format problems. => 1 1 1 0 0 1 0

disk format hard how my problems to

1 1 1 0 0 1 0

- униграммы
- биграммы
- триграммы
- n-граммы

Какова размерность?

Бинарная и многоклассовая классификация

- Входит и не входит
- Входит в один из нескольких
- Входит и не входит (входит в другой?)

Кластеризация и тематическое моделирование

- Попадает только в одну группу
- Попадает разные группы

Регрессия

- Прогнозирование
- Линейная
- Линейка и резинки
- Мягче или жестче - переобучение
- За какой период? Внешние факторы

Понижение размерности

- Garbage in, garbage out (GIGO)
- Корреляция

Рекомендации

Ищу сантехника.

- Мне посоветовали.
- Опросил много людей
- Реклама это хорошо?

Рекомендации

- Сбор данных
- Кластеризация документов
- Кластеризация пользователей
- Как определить что рекомендовать?
- Идеальные рекомендации или грязь полезна

ДЗ

86.116781;55.335492;2;4326901.00;54.00;7.00;5;5
79.901011;55.876427;2;2992422.00;46.00;6.00;3;5
86.409119;53.593361;2;3578989.00;54.00;10.00;3;17
85.279779;56.566278;3;11484737.00;94.00;25.00;10;10
79.786809;52.654771;4;7477349.00;139.00;24.00;5;12

Спасибо за внимание!